DOCUMENT RESUME

ED 331 498 IR 052 750

TITLE	Database Management: Building, Changing and Using Databases. Collected Papers and Abstracts of the Mid-Year Meeting of the American Society for Information Science (15th, Portland, Oregon, May 1986).
INSTITUTION	American Society for Information Science, Washington, D.C.
PUB DATE	May 86
NOTE	37p.; Microfiche containing papers from the conference not available from ERIC.
PUB TYPE	Collected Works - Conference Proceedings (021) -- Viewpoints (Opinion/Position Papers, Essays, etc.) (120) -- Reports - Descriptive (141)
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Abstracts; *Database Design; *Database Management Systems; *Databases; *Information Retrieval; Information Systems; *Microcomputers; *Optical Data Disks

ABSTRACT

This document contains abstracts of papers on database design and management which were presented at the 1986 mid-year meeting of the American Society for Information Science (ASIS). Topics considered include: knowledge representation in a bilingual art history database; proprietary database design; relational database design; in-house databases; alternate indexing schemes; Diselector, a diesel engine database; end user-created databases; database management system selection; the Department of Defense Gateway Information System; technical staff needed for database management system maintenance and enhancement; integrating database management into the library school curriculum; optical disk databases; remote database use and management; use of citation counts for faculty evaluation; development of a research and development tracking system; hypermapping; a biotechnology knowledge-based system; design and implementation of an automated correspondence program; database use in management; document and data retrieval; design of a menu-driven user interface for the ERIC database; library databases; and information retrieval system design. (EW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Databases

ED331498

asis

American Society for Information Science

DATABASE MANAGEMENT: BUILDING, CHANGING, AND USING DATABASES

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

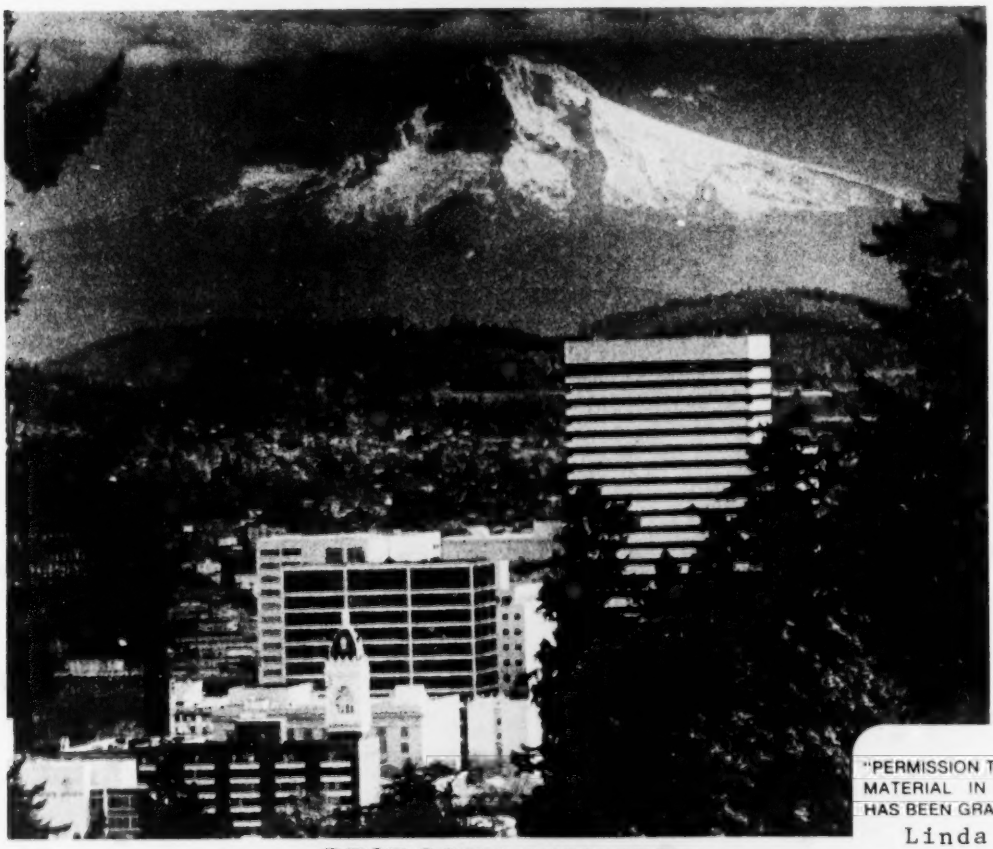
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Collected Papers and Abstracts
15th Mid-Year Meeting
Portland, Oregon
May 1986

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY
Linda Resnik

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."



BEST COPY AVAILABLE

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY
Linda Resnik

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

KNOWLEDGE STRUCTURE REPRESENTATION: PLANNING & DESIGN FOR A BILINGUAL FRENCH-ENGLISH ART HISTORY DATABASE

James D. Anderson
School of Communication, Information and Library Studies
Rutgers University
4 Huntington Street
New Brunswick, New Jersey 08903
January 1986

ABSTRACT

Information retrieval databases may be viewed as representations of knowledge in the fields they cover. Current practice in bibliographic database design and implementation tends to provide extensive structure and definition to elements of document description while knowledge description is frequently relegated to a few, relatively unstructured fields. This paper describes the design of a bibliographic database for art history which provides a structure and definition of the knowledge represented in its literature. Placed in a framework of fundamental facets and roles, a descriptor matrix is used to generate classified arrays and alphabetical contextual indexes in both French and English.

The J. Paul Getty Trust, which has funded this project, and the French Centre National de la Recherche Scientifique have agreed to merge the two major international abstracting and indexing services in Art History. The new jointly produced database will feature abstracts in Fre..ch or English, depending on their source, with complete indexes in both French and English.

Descriptors: Art history; Abstracting; Bibliography; Bilingual; Classification; English; French; Information retrieval; Indexing; Knowledge representation; Knowledge structure; Online.

DESIGN OF A HIGHLY SECURE MICROCOMPUTER-BASED SYSTEM FOR A PROPRIETARY DATABASE

Ruthann Bates,* Richard A. Roth,* Helen Welsh,**
and Barbara Wood**

* Aspen Systems Corporation
** Rohm and Haas Company

A self-contained microcomputer system, involving no telecommunications and installed in a secure facility, was designed to prepare a data base of highly proprietary chemical research reports. The system is based on two IBM PC XT's with associated printer, tape drive, and other equipment. It is installed in a secure room within a secure facility; all work is performed in that room. The system produces records in BRS OCDS format that are loaded on a mainframe computer for searching via BRS/Search. Proced res for handling the proprietary information, design considerations for the microcomputer configuration, security considerations for both the system and the facility, and results to date are discussed.

Descriptors: microcomputers; security; chemical information

FOUR APPROACHES TO RELATIONAL DATABASE DESIGN

Terrence A. Brooks

Abstract

Relational database design is considered in terms of microcomputer database management programs. The importance of effective database design and the lack of criteria for effective designs are discussed. Four possible approaches to relational database design are considered: the universal relation, normalization, synthetic designs, and representational redundancy.

Descriptors: Relational Database Management Systems; Database Design; Normalization

USE OF MICROCOMPUTERS FOR RETRIEVAL AND ANALYSIS OF BIOLOGICAL INFORMATION

Joseph P. Brown

Oncogen, 3005 First Avenue, Seattle, Washington
Department of Pathology, University of Washington
Medical School, Seattle, Washington

During the past decade two revolutions have been under way. The first has resulted from the introduction and widespread acceptance of the microcomputer. The second revolution has been in biology. Powerful methods for cloning and sequencing DNA molecules have resulted in the determination in thousands of genetic sequences, comprising tens of millions of nucleotides.

These genetic sequences have been compiled into databases, the most well known of which are GenBank (DNA sequences), and the Protein Identification Resource (PIR, amino acid sequences). These databases can be accessed on-line, either directly (PIR), through a third party such as Intelligenetics, or through a local mainframe or minicomputer. Unfriendly operating systems, a lack of portable software, and expense are major problems of working on-line.

Most of the computational tasks associated with DNA and protein sequence analysis do not require either the storage capacity or computational power of a mainframe or minicomputer. A thirty megabyte hard disk, which can be obtained for less than $1,000 can hold both the DNA and protein sequence libraries, and the most common operations, such as translation of a DNA sequence to amino acids, or searching for restriction enzyme sites take only seconds on an IBM PC. Sequences can be retrieved from the databases by an indexed keyword search in a few seconds, while a homology search may take ten minutes to compare a protein sequence against the PIR library.

The DNA and protein sequence databases are now being distributed on floppy discs or tapes that can be read by an IBM PC, and sequence analysis software is now available commercially. The use of microcomputers for genetic sequence retrieval and analysis is becoming more and more widespread, and indeed, indispensible.

Microcompters, DNA, Protein, Sequence database, Retrieval

THE DEVELOPMENT OF IN-HOUSE DATABASES AT R. J. REYNOLDS TOBACCO COMPANY

Helen S. Chung, Randy D. Ralph and Ronald E. Shackelford

R. J. Reynolds Tobacco Company
Bowman Gray Technical Center
Winston-Salem, N.C. 27102

ABSTRACT

An automated in-house information storage and retrieval system was developed at the Research and Development Library, R. J. Reynolds Tobacco Company. The goal of this automated information management system is to help R&D Library reduce redundant files and provide flexibility in online information retrieval. The library information system was designed and developed using MINISIS. MINISIS is a database management system supported on the Hewlett-Packard HP3000 minicomputers and was developed by the International Development Research Center of Ottawa, Canada. The selection of MINISIS software is discussed. This paper describes the development, design, structure and content of the R&D Library in-house databases and their applications.

DESCRIPTORS:

Information Storage and Retrieval, Library Automation, Information Management, Database Management System, Databases, MINISIS, ILO, Menu-Driver, User-Friendly Interface

Working Paper: Methodology for Comparative Evaluation of Alternative File Indexing Schemes

By: Donald Conroy, Jr.
Marketing Software Support Specialist
URICA Library Systems
McDonnell Douglas Computer Systems Company

Lawrence Woods
Manager, Product Development
URICA Library Systems
McDonnell Douglas Computer Systems Company

ABSTRACT:

As information databases have grown both in size and complexity, so has the need to develop effective and efficient file indexing schemes to facilitate user access to those databases. This paper describes a methodology being developed by the technical staff of the URICA Library System Development Center at McDonnell Douglas Computer Systems Company to evaluate alternate indexing schemes.

Issues to be considered in such evaluations are: coding algorithms, searching algorithms, file sizes and file retrieval times.

Schemes compared are such classical cases as Soundex, Davidson Consonant Code, Keyword Approximation, and other permutations. A major result of this effort is a methodology to compare current schemes with new potential index algorithms. Examples are included of actual computer test runs.

DESCRIPTORS:

Soundex, Davidson Consonant Code, Keyword Approximation, Algorithms, Searching, Library Automation

DEVELOPMENT OF DIESELECTOR
CASE STUDY OF A NEW KNOWLEDGE BASED PRODUCT

Dr. Marianne Cooper
Graduate School of Library and Information Studies
Queens College
of The City University of New York
Flushing, New York 11367

A novel combination of an off-the-shelf database manager, custom written computer programs, and a database has been prepared to locate and manipulate stored data to answer users' questions about diesel engines. It is now being marketed for use on microcomputers as DIESELECTOR. This paper presents the initial activities related to product definition, data acquisition, database creation and quality control, and programming, and illustrates the need for the cooperative efforts of various professional disciplines at different stages of the development of a value added knowledge based product.

* DIESELECTOR is a registered trademark
of Dynalytics Corp, Hicksville, New York.

Descriptors: Diesel Engines, DIESELECTOR, Dynalytics Corp.

COMPARING RETRIEVAL METHODS FOR AN END USER-CREATED DATABASE: A CASE STUDY.

Miriam B. Dunbar and Joyce Kho

Abstract:
Computer conferencing is being used increasingly for group communications. The resulting transcript consisting of electronic messages by the participants is unique compared to other kinds of documents. For the contents to be useful, some method of accessing is needed. This case study describes the human, intellectual effort of designing a controlled vocabulary and the use of a free-text information retrieval software. Some combination of machine and human indexing will provide the best access results.

Descriptors:
Indexing; Information Retrieval; Controlled Vocabulary; Free-text searching; Computer Conferencing.

THE ART OF DBMS ACQUISITION

Gordon M. Gerson, Sr.

This paper discusses several of the non-quantifiable aspects of Data Base Management System selection. Rationale for a methodology of DBMS are presented.

Descriptors: DBMS, macro-analysis, consultants, analysis, software selection, software evaluation

THE DEVELOPMENT OF MICROCOMPUTER DATABASES FOR ARCHIVAL APPLICATIONS

Anne J. Gilliland

Abstract: A discussion of current and potential applications for microcomputer-based database management systems within archival repositories, including the evaluation of software and the structure of an administrative database to accommodate the specialized nature of archival arrangement and description; use as a reference database; potential for online public access and for integration with library information systems; the uploading of data on holdings into bibliographic utilities; records management applications; and the generation of customized reports and indexes.

Archives
Bibliographic utilities
MARC format
Microcomputers
Online public access
Records management

FILE FORMAT CONSIDERATIONS IN MICROCOMPUTER DATABASE MANAGEMENT SYSTEM SELECTION

Lynn C. Heer and Thomas Kinney

ABSTRACT

File Format Considerations in Microcomputer Database Management System Selection

An important factor for consideration in selecting a microcomputer database management system (dbms) is the system's ability to exchange data files, including both the types of files which may be imported into the database and the file formats which may be exported for use in other programs and on other computers. The range of formats supported determines the user's ability to transfer data from and to other programs and other computer systems. This also impacts the development of custom programs for processing these data files. Common data file formats (DIF, SYLK, DBF, CSV, etc.) and three widelyused microcomputer information managers (dBaseIII, R:base 4000 and Lotus 123) are discussed in terms of data file format considerations in software selection. Various options exist for data file transfer and should be considered. A glossary of selected terminology related to microcomputer data file formats has been prepared for distribution.

DESCRIPTORS: Microcomputers; data structures; file structures; file formats; data interchange; data files; database management SDI, WKS, VC; SDF; DIF; SYLK; ASCII; CSV; Lotus 123;dBase; R:base 4000; data communication; data transfer; file transfer; data exchange

THE EVALUATION OF MAINFRAME DATABASE MANAGEMENT SYSTEMS

By
Charles Holmberg
Deputy Director - Data Processing
Stanislaus County Data Processing
1100 H Street, Room 1
Modesto, California 95354
(209) 571-6394

Stanislaus County Data Processing provides a wide range of computer services to over 35 departments in areas as diverse as Legal Justice, Financial Systems, and Social Services. In order to meet the growing information needs of these users, the County Data Processing Department adopted a philosophy of utilizing database software for the development of all new applications.

The paper I intend to present to the ASIS Mid-Year meeting, will detail the purpose, analysis and structure of the evaluation process of mainframe database software. Specific categories to be discussed are software interfaces, DBMS design, data dictionary, query facilities, report writers, recovery/restart, DBMS security, operational efficiency, installation considerations and training. Finally, the paper will discuss the scoring/ranking technique utilized to determine the overall ranking of the individual DBMS packages being evaluated.

Descriptors

- Database Management System
- Relational Database
- Software Evaluation
- Hardware/Software Compatibility
- Interface
- Design
- Data Dictionary
- Fourth Generation Language
- Security
- Questionnaire
- Documentation
- Training
- Scoring

THE DOD GATEWAY INFORMATION SYSTEM DIRECTORY OF RESOURCES

Carol E. Jacobson and Gladys A. Cotter

The Defense Technical Information Center (DTIC) is sponsoring development of a DoD Gateway Information System (DGIS) to provide online, streamlined methods for identifying, accessing, searching and analyzing data from heterogeneous databases of interest to the DoD community. At the core of the gateway system is a Directory of Resources which contains information on the content, scope and availability of selected databases. Presently, the Directory references over 430 DoD-sponsored R&D databases, but this figure is expected to grow to over 3,000 when coverage is expanded to include additional DoD, other federal, and commercial databases. The prototype Directory runs on a VAX 11/780 minicomputer using the INGRES database management system and is available to a select community of users. A critical feature is the menu-driven interface designed to assist end-users. This paper describes data collection, database design, implementation of the Directory and menu-driven interface, and future directions.

Online Directories, Database Management Systems, Gateways, Database Design, Requirements Analysis

ABSTRACT
American Society for Information Science
Mid-Year Meeting
May 11-14, 1986
Portland, Oregon

Contributed paper title: "Many Actors in One Play: The Differing Roles of Database Management, Staff, and Users."

Author: Darrell L. Jenkins
Director of Library Services
Morris Library
Southern Illinois University at Carbondale
Carbondale, IL 62901
618-453-2681

Descriptors: Planning and design, role of database administrator, politics of network/database administration; User interface role of administrator vis-a-vis role of users, multi-user environments

Abstract:

Much of the current literature regarding information networks and database management emphasizes the capabilities of the networks, their hardware and software requirements, specifications, internal indexes, and applications. An area usually given some attention is the requirements of the technical staff needed to maintain and enhance the system.

However, with a shared database comes different types of users, different user needs and expectations, and resulting political ramifications as these users interact among themselves and with the database administrator and staff. It is this political interaction that plays a major role in the acceptance, use, and success of an information network.

INTEGRATING DATA BASE MANAGEMENT CONCEPTS, PRINCIPLES, AND SKILLS INTO THE MLS CURRICULUM

by
Thomas R. Kochtanek, Ph.D.
Department of Information Science
University of Missouri-Columbia
Columbia, Missouri 65211

ABSTRACT

Data base management principles form the basis for a large number of automation applications in libraries and related information agencies. The coverage of these fundamental principles should be highly prioritized curricular objectives within the MLS masters degree program. Such fundamentals can be introduced early in the student's program and reinforced via application in various courses within the curriculum. Sophisticated microcomputer data base packages allow students hands on experience in designing, developing, and using scaled down versions of larger implementations.

A brief description/overview is given for several of the commercially available data base management packages which operate in the microcomputer environment and have potential for applications in the professional setting. A listing of both selection criteria for these packages and a listing and discussion of facts, concepts, and principles relating to data base management is provided. The paper concludes with a recommendation for a possible curricular design which would facilitate the transfer and retension of those data base fundamentals.

KEY WORDS

Data base managers
Data base management systems
Automation
Microcomputers
Education
Curriculum

DBMS AND BIBLIOGRAPHIC RETRIEVAL: A COMPARATIVE ANALYSIS

James Krikelas

and

Patrick Wenzel

ABSTRACT:

The authors report on a study investigating the merits of microcomputer-based database management systems (DBMSs) for bibliographic storage and retrieval. Initial efforts to establish a benchmark for comparative evaluation are discussed and the issue of "transporting" data from one system to another is covered. Characteristics of a range of systems including Finder, Sci-Mate, InMagic, and PC File III are described. A summary of a comparative analysis of two systems -- KnowledgeMan and dBase III -- is provided. Suggestions for an improved bibliographic DBMS system are presented.

DESCRIPTORS:

Database Management Systems (Microcomputer)
Information Retrieval
KnowledgeMan
dBase III
Software Evaluation

OPTICAL DISK DATABASE SERVERS

Clifford A. Lynch
Division of Library Automation
Office of the President and Universitywide Services
University of California
186 University Hall
Berkeley, CA 94720

Abstract

Optical storage technology holds great promise as a means of publishing large databases and as a means by which an installation can place large databases (plus installation-specific indexing) online. But the full potential of optical disk databases will be realized only when they are integrated into a network environment that facilitates ubiquitous (remote) access and resource sharing of costly databases.

The integration of optical disk databases into existing systems has been greatly complicated by the lack of standards in the area and by the tight coupling between the arrangement of the data on the optical medium and the retrieval software that accesses the data. The construction of optical disk database servers would define a fixed set of interfaces between the optical disk database "system" (e.g., the disk and its tightly coupled retrieval software) and the external world.

This paper focuses on the appropriate level of abstraction in developing such protocols. It examines several of the existing models for network disk servers and discusses the potential applicability of the Linked Systems Protocol to the design of optical disk database servers. Experimental protocols under development at the University of California's Divison of Library Automation are also briefly discussed.

DATABASE DESIGN FOR READ-ONLY MEDIA

Clifford A. Lynch and Edwin B. Brownrigg
Division of Library Automation
Office of the President and Universitywide Services
186 University Hall
Berkeley, California 94720

Abstract

Organizing a database for publication on read-only optical disk presents an unusual set of choices for the database designer. Cost tradeoffs in the optical disk environment differ radically from the more usual shared mainframe and magnetic disk environment. We present simple cost models that provide some insight into these differences. As part of this discussion, we also explore approaches useful in maintaining read-only databases on magnetic media.

The paper reviews a number of techniques appropriate for read-only databases in the optical disk environment. Since space is not at a premium for many applications, it is feasible to use data structures that provide very fast access at the expense of consuming more space than would be acceptable for most databases on magnetic disk. When the size of the database to be published approaches the capacity of the publication medium, on the other hand, powerful compression techniques can be applied both to the data proper and at the file organization level. Computing time expended during the mastering process can return a high payoff in the resulting database. This processing time then can be amortized across large numbers of copies of the database, which are usable on relatively low-cost hardware.

MAPPING INFORMATION HIGHWAYS FOR THE NORTHWEST
(remote access; networks, directories)

By Ruth-Ellen Miller

Managing and using remote databases requires having access to them from remote sites. In an effort to improve such access for information users in the region, the Fred Meyer Charitable Trust Library and Information Resources for the Northwest program funded a project to create a directory of the channels by which information and data could be transmitted through the region.

The published Directory (Murr, Williams and Miller, Information Highways, Hypermap, 1985) was designed as a relational database in printed atlas form. Using computer-generated graphics, the project team condensed many volumes of information into one. The contents are divided into 3 sections: the technologies (what they are and how they work), the publicly accessible systems and "highways" in the region (where they are and how to gain access to them), and the basics of designing one's own systems. This presentation will provide an overview of the contents of the directory and the process for developing it.

CITATION COUNTS FOR FACULTY EVALUATION:

ONLINE VS. MANUAL COMPILATION TECHNIQUES

Pamela A. Mofjeld and Nancy G. Blase

Natural Sciences Library, University of Washington

Seattle, Washington 98195

Faculty have begun to utilize citation counts for promotion, tenure and merit decisions. In this paper, the authors describe the technique they have developed using the online databases SCISEARCH and SOCIAL SCISEARCH for compiling citation counts. The technique takes into account senior-authored and co-authored publications and self-citations. Evaluation factors to consider, methodological problems of using citation indexes, printed and computerized, for the citation counts, and comparisons between faculty working in the scientific and social science fields are covered.

DESCRIPTORS: SCISEARCH, SOCIAL SCISEARCH, citation counts, faculty evaluations

THE DEVELOPMENT OF AN R&D INFORMATION TRACKING SYSTEM IN THE PHARMACEUTICAL FIELD

Lawrence G. Mondschein
Janssen Pharmaceutica Inc.

Abstract. This paper describes the evolution of an integrated research information system at Janssen Pharmaceutica Inc., a member of the Johnson & Johnson Family of Companies. The Janssen Research Information System (JANRIS) monitors the progress of clinical research investigations from grant request approval through submission to the Food and Drug Administration (FDA) in support of a New Drug Application (NDA). It is operating on the IBM System 38 mainframe computer.

The system was initially developed in 1983, using List Processing on the Wang Office Information System (OIS). In early 1984, it was converted to the IBM System 38 due to problems with storage and response time on the OIS. This was in addition to the proposed enhancements to the existing system that could not be supported using List Processing.

JANRIS, which is maintained by the R&D Division, is actually a composite of four separate database systems, each individually under the responsibility of the R&D or Finance Divisions. Data on the regulatory status, patient recruitment, site visits, and the progress being made on the completion of the statistical and clinical research reports are included in the protocol monitoring section of the system. Financial data are retrieved via a direct link to the general ledger and accounts payable systems. Two types of reports are generated: summary reports monitor the progress of each drug trial, and alerting reports are reminders to the staff of approaching deadlines. These reports are distributed to all divisions within Janssen Pharmaceutica Inc. and to selected Janssen affiliates in other countries. The summaries have been found to be a valuable tool used by the research staff, not only to keep informed on the status of all our drug submissions, but as a reminder of approaching deadlines.

The future direction of JANRIS is on two fronts: 1. To download information from the system to the microcomputer for further analysis using a variety of commercially available software packages, and 2. R&D workload forecasting utilizing the computer generated reports to assist management with long range planning. The ultimate goal of JANRIS is to facilitate the channels of communication whereby information can best be disseminated and utilized. This will assist in the projection of realistic submission dates that can assure the filing of quality drug applications to the FDA in the shortest period of time.

Descriptors. Research tracking systems, clinical information systems, proprietary information, database management.

HYPERMAPPING: EXAMPLES OF DATABASE CONVERGENCE AND CONNECTIONS TO FORM GRAPHICS AND IMAGE NETWORKS

L. E. Murr and J. B. Williams

Oregon Graduate Center
Beaverton, OR 97006 and
Oregon Health Sciences University
Portland, OR 97201

The inability to deal with massive quantities of data in numerous generic areas, particularly involving technologies, demands new approaches and methodologies. In assimilating large volumes of data and utilizing databases, we tend to generalize concepts by developing images or pictures. Indeed, a powerful feature of human intellect is the ability to transform textual descriptions and textual data into images which summarize complex concepts. The embodiment of this concept is perhaps epitomized in the commonly heard phrase "a picture is worth a thousand words." When properly formatted, however, pictures can be connected into networks eliciting concepts which embody tens of thousands of words.

This paper illustrates some methodologies for connecting networks of images and graphic information including text blocks. We call these graphic-text networks hypermaps. Applications utilizing hypermaps to conceptualize complex ideas or to do conceptual planning based upon the creation of pictures of existing technologies (especially telecommunications/information delivery technologies) we call hypermapping. We present examples of hypermapping which involve the creation of maps of public telecommunications and information delivery networks and associated technologies in the Pacific Northwest: Alaska, Washington, Oregon, Idaho and Montana. We use these maps to create numerous conceptual plans for developing new networks, and to simplify the extent of information delivery technologies which exist today. We also illustrate the use of hypermapping to deal with complex information and knowledge development in nonserial formats which allow the learner to rapidly assimilate information by creating pictures.

DESCRIPTORS

Communication Technologies, Communication Techniques, Pacific Northwest Public Information Nets, Networks, Network planning, Telecommunications, Teleconferencing - Audio, Video, Computer. Databases. Public Libraries in the Pacific Northwest. Systems Planning. Knowledge Development.

TOPSY
A MICROCOMPUTER BASED EXPERT FOR COMMERCIAL INFORMATION MANAGEMENT AND USE IN BIOTECHNOLOGY

JAMES R MURRAY,PhD

POLICY RESEARCH CORPORATION INTERNATIONAL
CHICAGO, ILLINOIS
MARCH 1986

ABSTRACT

There are really only a few items that need be believed to make this effort worthy of pursuit.

BIOTECHNOLOGY IS EXTRAORDINARILY DIVERSE
IT IS SUBJECT TO SIGNIFICANT REARRANGEMENT AS NEW KNOWLEDGE
MONEY IS AS IMPORTANT AS SCIENCE
INFORMATION MATTERS
MURPHY'S LAW
COMPUTERS DO SOME THINGS CHEAPER THAN PEOPLE

Since early 1984, we have been developing/revising/implementing a computer based system to support strategic activities in biotechnology. The ultimate goal for this system, TOPSY, is that it perform the duties of an Expert System for those wishing to make a few dollars in biotechnology. Uses include:

1-SHOW AND TELL: a simple organized rendering of knowledge about activities in queried areas, e.g. firms, products, technologies etc.

2-VENTURE: an analysis of real or proposed firms in terms of the system's evaluation of their economic value, prognosis, etc.

3-OPPORTUNITY: areas of activity which appear to be under-exploited in terms of both economics and the science/technology.

4-APPLY: an analysis of existing technology with respect to a particular application.

5-PRODUCT: an analysis of particular product activities.

Descriptors: Biotechnology, Knowledge Based Systems

"OPTICAL DISK SYSTEMS - TODAY & TOMORROW"

Allen W. Paschal

The following paper will give you some background on DATATEK, DATATEK's mission and how DATATEK is attacking the problems of storage and communications, and specifically, how optical disk technology will change our industry.

dBASE + TEX = BEAUTIFUL LETTERS

M. Rabindranathan
Sally Lawler

Mathematical Reviews
P. O. Box 8604
Ann Arbor, Michigan 48107

Abstract

We describe the design and implementation of an automated correspondence program suitable for use by a library acquisitions department. The program is designed to achieve a high degree of modularity and flexibility. The discussion in this paper emphasizes the design aspects of the system. Also included is a section on hardware and software requirements.

Descriptors: Automated letter generation, computer typesetting, archiving, formatting, microcomputer database, dBASE, TEX.

Abstract

Database Design for Environmental Scanning

Eugene J. Rathswohl, PhD
School of Business Administration
University of San Diego
Alcala Park
San Diego, CA 92110
(619) 260-4861

This paper describes the requirements of a database design to support the information search activities involved in the environmental scanning phase of strategic management.

Information search can be considered a decision problem. Research tends to support the hypothesis that search is usually limited to the most readily available and familiar sources of information. These decisions to use certain information sources are frequently not conscious choices where alternative sources are explicitly considered. Research also tends to support the hypothesis that getting information faster is more important than getting it absolutely exact. What is required by managers developing strategic plans is the ability to quickly 'map' the information sources relevant to the critical environmental issues involved.

This paper discusses a 'knowledge representation' (database) framework being developed in a microcomputer environment to help management students search more effectively through an appropriate set of information sources relevant to environmental scanning tasks. The paper describes this work as a basis of an expert system to support environmental scanning activities.

Descriptors: Information search; environmental scanning; decision support

DATABASE ADMINISTRATION IN A MATRIX ORGANIZATION: PLAYING THE GAME AND COACHING THE TEAM

L. M. Roseberry

In a matrix organization, the database administrator plays on the program team and coaches the project team. Since the program structure cuts across line organization, program allegiancies compete for attention, energy and loyalty with line administration. This is a valid conflict. Future funding and growth depend upon program (matrix) attention to coaching the members of the line team. How well efforts are balanced between these conflicting demands determines the success rate of a database administrator in a matrix environment.

Line Management
Program Management
Funding
Matrix Organization
Database Administration
Resource Allocation

PARALLEL APPROACHES TO DOCUMENT AND DATA RETRIEVAL

Gerard Salton*

Abstract:

The conventional approach in designing integrated systems for document and data retrieval consists in extending a standard data base management system by adding special structures and operations that accommodate variable length text strings. An alternative approach might be to use a conventional document retrieval procedures and add certain data management facilities.

In the present note a different strategy is outlined which consists in using operations that are simultaneously applicable to both document and data retrieval. The proposed integrated system manipulates objects with mixed identifiers, including both objective attribute values, as well as subjective content descriptors. File organizations are used that permit both browsing as required in document processing, as well as objective file processing normally used in data base management. Certain features of the integrated design are outlined, and experimental data are included.

Descriptors: data retrieval, document retrieval, integrated retrieval system, complex records, composite documents, content identifiers, attribute-value, record relationship indicators, extended Boolean retrieval, hierarchical clustering, cluster searching

---

Department of Computer Science, Cornell University, Ithaca, NY 14853.
This study was supported in part by the National Science Foundation under grant IST 83-16166.

Molecular Biology Computing Using Databases of Biological Sequences

Dennis H. Smith
BIONET[tm] Resource Manager
IntelliGenetics
1975 El Camino Real West
Mountain View, CA 94040

Databases of biological sequences, both nucleic acids and proteins, are becoming an indispensable tool for the conduct of research in molecular biology. Such databases are used to compare and contrast newly indentified sequences with those already in these repositories, as scientists attempt to determine details of the structure and function of new genes and the proteins which they encode.

The use of such databases by scientists differs in three important, interrelated ways from conventional use of databases. First, the actual sequences contain a great deal of information on structure and function, but the sequences themselves cannot easily be characterized or indexed for use in conventional DBMS's. Second, because of the first difference, programmatic access to the sequences is required, so that various analytical programs can be used to perform the necessary retrievals and comparisons. Third, because these procedures require step by step interpretation of the results, information intermediaries generally play no role in accessing the databases; such access is performed by the scientists themselves, using interactive computer programs. The implications of these differences for systems of the future will be discussed.

Finally, the roles of existing regional and national resources for collection and analysis of biological sequence data will be described. These resources include the BIONET National Computer Resource for Molecular Biology, the GenBank Genetic Sequence Databank, The Protein Identification Resource, the Molecular Biology Computer Research Resource and the Molecular Biology Information Resource.

Descriptors

computer resource, genetic sequences, genetic sequence databank, BIONET Resource, GenBank, Protein Identification Resource, nucleic acid sequences, protein sequences, sequence analysis programs

A MENU-DRIVEN INTERFACE FOR INTERACTIVE QUERY REFINEMENT IN BIBLIOGRAPHIC DATABASE SEARCHING

Oren Sreebny
SDC Information Services
2525 Colorado Avenue
Santa Monica, CA 90406

Describes the design and exploratory test of a menu-driven user interface to the ERIC database. The interface design was based on Fidel's model of moves made by experienced online searchers. The interface presents menus which offer various options for enlarging or reducing the absolute size of the initially retrieved document set. Fourteen graduate students searched the same information request. Nine of the searchers used the menu-driven interface, while five used the normal system command language. The searchers using the menu-driven interface used fewer search keys and commands and used more commands which utilized complex features of the database.

User Interfaces; Information Retrieval; End-User Searching; Exploratory Studies; Bibliographic Databases; Online Searching; Interactive Systems; Menu-Driven Software; Search Strategies; Microcomputer Software; Intermediary Systems.

A STRUCTURAL RELATIONAL MODEL FOR THE DESIGN OF INFORMATION SYSTEMS USING DATA BASE MANAGEMENT SYSTEMS

Dr. Renee Tjoumas
Assistant Professor
Queens College of the City University of New York

Noel J. Squitieri
Business Systems Analyst
American Bankers Association

Abstract:

The American Bankers Association (ABA) has recently decided to replace its Digital Equipment Corporation Model 10 - KL mainframe computer and Datapoint Local Area Network with an IBM 4381 Group II mainframe and IBM 3270 PC AT desktop computers.

Over the past ten years the Information Systems Division has, using OLIVER - a Data Base Management System and programming language - developed a comprehensive information system providing such business functions as Accounts Payable, Accounts Receivable, Conference Management, Order Processing, Publications and Printing, and Subscriptions Management. These systems, and others, must be moved into the new IBM computer environment.

In preparation for the transfer of functional system components to the new computer a systems design model has been developed at the ABA by Frank Schlier, Director of Information Systems and Dr James Rader, Associate Director of Information Systems, which uses the capabilities of relational data base management systems. The model provides a framework for the design of an information oriented corporate data base which can be shared by all departments within a company. The model is presented and its use is discussed.

Descriptors:

DBMS, database, database design, relational

PROBLEM-SOLVING USING THE SCIENTIFIC METHOD: A SIMPLIFIED LOGICAL MODEL WITH APPLICATIONS FOR INFORMATION SPECIALISTS

Rodes Trautman
Graduate Library School, University of Arizona
1515 E. First Street, Tucson, AZ 85719

The most celebrated problem-solving procedure for essentially any question or assignment is the scientific method. It is commonly described by a linear listing of several (three to nine) steps. The steps are to be repeated in a recursive or cyclical manner to solve ever widening applications or to achieve ever deepening analyses.

Procedures involving decisions and recursion are more easily described -- and grasped -- if displayed graphically in a simplifed model with a small number of logical components. Such a generic model for the scientific method is proposed that has just four steps -- called DEFINE, COLLECT, ANALYZE, and PREDICT -- around a collection of exemplars and guidelines -- called the PARADIGM.

Applications for librarians and information specialists will be discussed because models are *good*, *reasonable*, or *bad* according to our experience in using them. These examples include writing a library research paper, reporting empirical research in a primary journal article or patent application, and the set of 19 primary verb commands proposed by the National Information Standards Organization Z39-G committee for a common database retrieval language.

Modelling, paradigm, deduction, induction, primary literature, reference process, NISO Z39, common command language

SUBJECT ANALYSIS OF THE OCLC DATABASE

Diane Vizine-Goetz and Edward T. O'Neill
OCLC Online Computer Library Center, Inc.
Dublin, Ohio

The OCLC Online Union Catalog was analyzed to obtain statistics on the distribution of monograph records by subject. To accomplish this, about seven million records were processed against the major Library of Congress classes and the 490 National Shelflist Subject Categories. Data on the distribution of monograph records over subjects should be useful for selecting database subsets, for estimating the expected growth of a subject collection, or for comparing local collections with the OCLC database.

OCLC Online Union Catalog
Library of Congress Classification System
National Shelflist Subject Categories
Subject Collections

A DATABASE IN TRANSITION: THE EVOLUTION OF THE WINDWARD COMMUNITY COLLEGE LIBRARY HAWAII PAMPHLET FILE DATABASE

DeEtta Wilson, Martha Chantiny

The stages of development of the Hawaii Pamphlet File (HPF) database represent in microcosm the evolution, uses and power of database management, illustrating the transition from mainframe mystery to micro accessibility. The HPF is a subject arranged file of reports, pamphlets, brochures used by faculty and students at Windward Community College (WCC) in Kaneohe, Hawaii. An automated index providing access by subject, title and publishing agency was created in 1979. A computer science student wrote programs designed to run on the IBM 360 mainframe at the University of Hawaii at Manoa Computing Center. With use, certain problems emerged in maintaining the index on the mainframe computer. It was hoped that transfering the functions of the HPF index to WCC Library's IBM PC-XT would provide better control, easier access, and expanded options for use of the database. Evaluation of the functions and performance of the existing programs; description of an "ideal" redesigned, improved HPF database; examination and evaluation of two microcomputer database management software packages; and refining, creation and testing of a prototype database using Inmagic software are described.

Database management; Information storage and retrieval systems; Library collections, Technological advancement, Inmagic, IBM PC

PAINLESS INFORMATION RETRIEVAL: THE EFFECT OF GOOD USER INTERFACE ON DESIGN

by Lucy Anne Wozny

The Drexel Disk is a Macintosh microcomputer-based information retrieval system about Drexel University's Microcomputing Program and its relationship to the University. The disk contains information on microcomputing facilities, policies, programs, and short cuts and tips for the software most used by Drexel students, in both textual and graphical formats. The Macintosh user interface offers new opportunities to communicate with the user. It fosters a graphical and flexible interface for the system, always providing the user with three methods of access to the information: a graphical access through pictures representing topics, a textual access through menus with topic headings, and a direct textual access through an interactive index. When the user uses the graphical representation of the topics, he is stepping through a standard tree structure. Through menus within each topic, the user can see other information within the topic or return to the graphical structure. An interactive index to the Disk creates network-type access; users can jump to specific topics on the disk and automatically return to the index, use menus, or revert to the tree structure. The ability to quit the disk entirely is active at all times. Novices and experts to the Macintosh seem to be comfortable with the apparent flexibility of the retrieval system. Students and others shy of computers like the idea of clicking on a picture and getting information. Experts can go directly to the index and retrieve exactly what they need.

Information Retrieval System Design
User Interface
Microcomputers